SMIL: THE NEW MULTIMEDIA DOCUMENT STANDARD OF THE W3C

Nabil Layaïda and Ahmed Karmouch
Multimedia Information Research Laboratory
University of Ottawa, Department of Electrical and Computer Engineering
161 Louis Pasteur, Ottawa, ON, Canada, K1N 6N5. Tel: (613) 562-5800x6123
E-mail: Nabil.Layaida@inrialpes.fr

1 INTRODUCTION

The important Web growth and development have provided a unique opportunity for innovation, particularly in the fields related to Web infrastructure. New document standards like HTML 4.0 and XML, style sheet languages like CSS and XSL and network protocols like HTTP-NG and RTSP have been proposed to tackle unsolved problems at different levels in the original Web design. The Web evolved from very simple to more complete and complex environments that we use nowadays. If these later advances can be considered mainly as enhancements, multimedia presentations, which brings a newer vision to the Web, was not initially addressed. Web documents have been designed to be static. Their extension to coordinated dynamic behavior required for time sensitive data such as audio and video was not a key issue.

Parallel to the Web development, research in multimedia during the last decade has been very active. A large number of papers were published on synchronization both at the authoring and presentation levels. Document standards were also proposed. HyTime, MHEG and PREMO are good examples of such standardization effort. Unlike the Web, both experimental tools and document standards failed to have a real impact on commercial products like Macromedia Director or apple's Quicktime format. Standards were blamed to be too complex while research models and prototypes are still considered to be insufficiently mature.

In 1997, the Multimedia Synchronization Working Group called SYMM was formed within the World-Wide-Web consortium to come up with a proposal. A simple document format seemed to be a reasonable starting point to ease experimentation and at the same time to keep the door open for future extensions. A document format, termed *Synchronized Multimedia Integration Language* (SMIL for short) has been produced. An important key requirement for SMIL was to use available Web technology such as XML while promoting synchronization as a primary functionality. The milestone of the SYMM Group was a round started by the development of SMIL version 1, followed by an interoperability test to ensure its implementation by several research and industrial institutions. The goal was to restore a development and an enhancement cycle similar to the one that stimulated the Web development and growth.

The Multimedia Information Research Laboratory (MIRL) and Opera at INRIA (Institut National de Recherches en Informatique et Automatique) was involved in multimedia research work since its inception. We brought significant contributions to the SYMM working group since the initial call for interest launched in early 1996. Our vision was based on the experience gained in both projects in the design of Authoring and presentation tools (see Madeus [Lay97], MediaDoc [Kha95]). Our research work and prototypes served as a proof of concept implementation during the SMIL format specification.

This paper presents current achievements in the field of multimedia document specifications over the Internet within the W3C. It covers important requirements and choices that guided SMIL development. Particularly, we present the key features of SMIL and we discuss next generations of SMIL extensions.

2 MULTIMEDIA AUTHORING AND PRESENTATION

Many analysts predict that the Web will be turned into a distribution system for both interactive and continuous multimedia or *synchronized multimedia content*. A typical example of synchronized multimedia is a presentation created with tools like Microsoft's Powerpoint or Macromedia's Director. Most common examples can be found on CD-ROM products like training courses, virtual galleries, or presentations enhanced by text, images, and video. For instance, in a guided tour, the screen shows a sequence of different images, text, and graphs, which are played in parallel with an audio stream.

Many of today's television programs are also good examples of multimedia presentations and could be produced by using techniques found in multimedia CD-ROMs authoring. Consider a television news presentation: many parts of the television screen contain usually text, images, and graphics. These static elements are generally used to enhance the presentation and could be sent as separate data, together with a schedule that determines the timing for when these entities should be displayed on the screen. This may be much less demanding in terms of bandwidth than sending the same content as full motion video.

Several audio and video broadcasts are already sent through the Internet. Furthermore, providers of synchronized multimedia content are increasingly interested in using the Web as a medium as it becomes more and more economically attractive. First, the Web significantly facilitates the maintenance of the content. Therefore, up-to-date information can be offered via the Web almost instantly. Moreover, distribution of content via the Web is more cost effective than distributing CD-ROMs. With the forthcoming high-speed networks, it is likely that in the near future Web technology will become the ideal infrastructure for the content that is disseminated on CD-ROMs today.

Nevertheless, the domain of hypermedia documents imposes many challenges to the application designers. Particularly, the document creation and presentation are more complex than in traditional documents because of the dynamic nature of multimedia documents. Objects such as audio and video have inherent temporal dimension and authors can artificially add synchronization within a multimedia document and between these objects. The temporal dimension reflecting the ordering of media objects in a multimedia document is called a temporal scenario.

A scenario is based on streaming media presentations, which follows a given temporal ordering. For example, a video presentation on the Web contains two separate files, a video stream plus an audio soundtrack, that require coordination of how long each media stream plays and how the media streams relate to each other. A multimedia document allows specification of how and when the streams are played back according to a common timeline. For example, a 2-second delay can be easily set for the audio soundtrack without changing the encoded audio file. Thus, the video file can start playing first, and then 2 seconds later, the audio track starts to play.

The rest of this paper is organized as follows. In the following section, we mainly outline the representation and main advantages of SMIL compared to other document formats. Section 4 presents the document model and syntactic issues related to SMIL definition. It reflects the agreement reached for the requirements included in the first draft of the specification. It covers the design principles and an overview of the presentation scheduling based on the temporal information contained in the document. Important features of SMIL used as building blocks for a multimedia presentation are extensively presented together with simple examples. Section 5 concludes this paper and gives some hints for future extensions and enhancements of next generations of SMIL.

3 SMIL: toward a new standard for multimedia documents

SMIL (Synchronized Multimedia Integrated Language) is an open World Wide Web Consortium (W3C) Proposed Recommendation for the stylistic layout of multimedia presentations. SMIL defines the mechanism that authors can use to compose a multimedia presentation in combining audio, video, text, graphics. It also defines precisely where on the screen and when over time these media are presented to the viewer in synchronized manner.

Compared to other document formats, SMIL provides several advantages such as:

- **XML-compliant, widely supported format**. SMIL has the potential to become the common standard of Web-based multimedia. It allows content creators to choreograph the most dynamic, highly interactive streaming media presentations on any client-server platform with SMIL-support.
- **Powerful timing and control of multimedia presentations.** SMIL layout and synchronization tags define when and where multiple media elements can play within a multimedia Web broadcast. With SMIL, Web authors can easily control the presentation timeline.
- **Simple and easy-to-learn**. A SMIL file is a simple text file that can be generated using any text editor and the format is based only on a limited set of tags.
- **Scalable media streams.** SMIL supports multiple media presentation choices for different bandwidths providing the best viewing conditions at a given bandwidth.
- **Extensible media support**. SMIL supports multiple numbers of media clips and media types.
- **Flexible**. SMIL allows different language options for media files with different bit-rates and different screen resolutions.
- **Customized presentations.** Because SMIL is a simple text file, one can automatically generate different presentations based on preferences recorded by the user's browser.

SMIL was created and developed by the W3C SMIL Working Group, which includes key industry contributors such as AT&T/Bell Labs, Digital Equipment, Microsoft, Netscape Communications, Philips and RealNetworks. It involves also important research institutions like INRIA, CWI and Columbia University. At the time of this writing, SMIL has passed to the phase of proposed recommendation from a status of working draft. This means that in the next few months, the SMIL version 1.0 may become a recommendation if members votes are favorable. A recommendation is the equivalent to a recognized Standard in other organizations like ISO. This step is likely to happen as several companies have already announced their support for the proposal.

4 DESIGN PRINCIPLES AND FORMAT ENCODING

One of the most important goals of SMIL is to make multimedia technology more open in terms of specification. So far, Multimedia documents used in existing tools are very heterogeneous and the applications that handle them are very different and incompatible. The concept of structured document intends to homogeneous processing in hypermedia applications. For this reason, early decisions were made to use existing technologies as building blocks for SMIL.

Structuring has been used for several years in electronic publishing and has been recently brought to the Web documents through XML. The scope of XML is beyond the specific uses found in layout oriented mark-up as in HTML. It provides a clean separation between information on the Web like documents and objects and the tools that handle them. Such information is termed multimedia documents that exist in various formats usually proprietary like quicktime or real-audio. Although XML is extensible in essence, one of the important decisions that have been made for SMIL documents was to use two notations:

- An **XML Document Type Definition** (DTD).

- An augmented **Backus Naur Form** (BNF).

The key issue in using XML (Extensible Markup Language) is to provide more powerful tools to describe the document content. XML retains the SGML advantages of extensibility, structure, and validation in a language that is designed to be easier to learn, to use and to implement than the full SGML. Furthermore, XML differs from HTML in three major respects:

1. Information providers can define new tag and attribute names as needed. In SMIL new tags are required to describe synchronization.
2. Document structures can be nested to any level of complexity.
3. Any XML document can contain an optional description of its grammar for use by applications that need to perform structural validation. For SMIL documents such a tool is available at CWI and another is under development at the MIRL Laboratory.

While XML fits the requirements for a better description of the document organization and structure, SMIL documents requires more control on the syntax of the attribute values. Such attributes allow the description of important information on temporal synchronization of different elements in a document such as duration, begin and end instants, etc. In practice, attributes are only seen as character strings at the level of XML, but they play a much more important role in SMIL as they provide the basic information for temporal synchronization. For this reason a more precise description and control of these attributes through BNF is included in the SMIL specification.

4.1 Document Organization

At the document level, the structure reflects the temporal organization. That is synchronization is applied at the level of composite elements to all its children in the tree. This has been preferred to a logical organization because the goal was to focus on the temporal aspects. Despite the fact that a logical organization provides a cleaner separation between different dimensions of the document (ie. logical, spatial, temporal and hypermedia [Lay97]).

The structure of a SMIL document is composed of two main parts each representing information at two levels. First the **head** element contains information about the SMIL document. Such information is used to define properties of a document (e.g., author, expiration date, a list of key words, etc.) and assign values to those properties. Properties are described by "meta" elements and each of them specifies a single property/value pair. The head contains also the spatial layout description. Second, the body element that contains three building blocks:

1. **The temporal scenario** which is an encoding of the document temporal structure. This structure is based on the description of the temporal relations that are set between different elements in the document.
2. **Alternate descriptions** which covers the means offered to the author to choose between different portions depending on system and user defined attributes.
3. **Linking descriptions** which relate different documents or parts of document over the temporal dimension.

4.2 Scenario Specification

One of the popular methods to create a multimedia presentation is to use languages like Dynamic HTML or Document Object Model. In this case, programming provides the necessary support for temporal synchronization: objects are provided with methods (Java scripts) that triggers user-defined procedures, which in turn can trigger other objects in the presentation leading to a coordinated behavior. Programs are in-lined with Web documents and both provide a self-contained multimedia presentation. While this method have the advantage of working on already available technology (i.e. Java scripts enabled browsers), it suffers however from several drawbacks: multimedia authoring relies more on programming skills, which makes them inaccessible to most of the Web users. Furthermore, the stability of the document representation cannot be guaranteed as programming languages evolves rapidly. Finally, imperative languages such as scripts do hide the temporal information by scattering them in several procedures. This tends to make the document manipulation and maintenance a very tedious task, as scenarios become less readable. Instead, SMIL format is declarative. This means that authors have to specify the temporal ordering simply by describing the media objects contained in the document and the relations between them. The presentation system is responsible for achieving the synchronization between different objects based on these relations. Synchronization in SMIL is based on few simple temporal operators presented in the following sections.

4.2.1 The parallel element

The parallel element, tagged "par" is used to express that several synchronization elements occur in parallel. That means, the children of a "par" element can have arbitrary temporal ordering. The "par" element can have the following attributes:

Begin

This attribute specifies the time for the explicit begin of an element. The attribute can contain the following two types of values:

Delay-value

A delay value is a clock-value measuring presentation time. Presentation time advances at the speed of the presentation. It can be stopped, decreased or increased either by user actions, or by the player itself.

The semantics of a delay value depend on the element's first ancestor that is a ynchronization element (i.e. ancestors that are anchors ie. "a" elements or "switch" elements are ignored):

- If this ancestor is a "par" element, the value defines a delay from the effective begin of that element.
- If this ancestor is a "seq" element, the value defines a delay from the effective end of the first lexical predecessor that is a synchronization element.

Event-value

The element begins when a certain event occurs. Its value is another element event (begin, end or a delay from these instants). The element generating the event must be "in scope". The set of "in scope" element is determined as follows: all children from the element's first ancestor that is a synchronization element are included. All the children of all "a" elements are included, unless they are "switch" elements. The resulting set defines the "in scope" elements.

```
<par>
  <audio id="a" begin="6s" ... />
  <img begin="id(a)(4s)" ... />
   ...
</par>
```

In the following example, we outline two different uses of the parallel element. The audio element starts 6 second after the start of the par element (the triggering of its begin instant) while the image element is delayed 4 seconds from the start of the audio element. This relative positioning allows more flexibility for the authors as changes in the audio timing does not require changes for the image. While relative positioning is convenient for multimedia authoring it requires the tools to ensure that no inconsistencies are introduced. Obvious cases are cycles but they can be easily detected because the scope of relative positioning is restricted to in-scope elements. Other cases are more difficult to detect and further details of consistency checking can be found in [Lay96].

4.2.2 The sequential element

The children of a "seq" element form a temporal sequence. This means that every element in the sequence must end before the next element in the sequence can be started. Furthermore, delays can be inserted between elements. These delays refer implicitly to the previous sibling end instant

```
<seq>
  <audio .../>
  <audio begin="5s" .../>
</seq>
```

in the sequence. This restriction prevents the introduction of concurrency between elements when events are used. An example of delayed sequences is given in the following scenario:

4.2.3 Hard and Soft Synchronization

The accuracy of synchronization between the children in a parallel group is implementation-dependent. Take the example of synchronization in case of playback delays, i.e. the behavior when the "par" element contains two or more continuous media types such as audio or video, and one of them experiences a delay. A player can show the following synchronization behaviors:

- **Hard synchronization**: The player synchronizes the children in the "par" element to a common clock.
- **Soft synchronization:** Each child of the "par" element has its own clock, which runs independently of the clocks of other children in the "par" element.

4.3 Media Objects

One of the most important aspects of SMIL is the integration of Media objects. As we mentioned earlier, SMIL documents reflects mostly the structure of a scenario and information describing media objects. Except for text, a SMIL file does not contain the content of the objects but only references or locators that allow the content retrieval at presentation time.

The media object elements allow the inclusion of media objects into a SMIL presentation. Media objects can be included either by reference (using a URI), or they can be written inline (text). In the latter case, they are included as CDATA in the content of a media object element.

There are two types of media objects: media objects with an intrinsic duration (e.g. video, audio) (also called "continuous media"), and media objects without intrinsic duration (e.g. text, image) (also called "discrete media"). Anchors and links can be attached to visual media objects, i.e. media objects rendered on a visual abstract rendering surface.

When playing back a media object, the player must not derive the exact type of the media object from the name of the media object element. Instead, it must rely solely on other sources about the type, such as type information contained in the "type" attribute, or the type information communicated by the server or the operating system. Authors, however, should make sure that the group describing the media object (animation, audio, img, video, text or textstream) is reflected in the element name. This is in order to increase the readability of the SMIL document. When in doubt about the group of a media object, authors should use the generic "ref" element. Basic media objects are described using the following tags: The `ref`, `animation`, `audio`, `img`, `video`, `text` and `textstream`.

4.4 Layout Specification

In contrast with text oriented documents having only a spatial dimension, multimedia documents have a temporal dimension. An SMIL document may contain a layout section that determines the placement of the presentation components in non-temporal dimensions. The layout section may contain several alternative layout elements embedded within a switch element. This can be used for example to describe the document layout using different layout languages. The current specification defines only a basic layout language for SMIL.

SMIL basic layout is a subset of the visual rendering model and the formatting properties defined by the CSS2 specification [CSS98]. SMIL basic layout only controls the layout of a subset of SMIL elements referred to as *positionable elements*. In SMIL 1.0, only basic media object elements are positionable. It is illegal to use SMIL basic layout for elements that are not positionable, for example entire portions of parallel or sequential parts of the scenario.

SMIL basic layout makes the following restrictions to the CSS2 visual rendering model:

- All positionable elements are *block-level* elements (i.e. the value of the "display" property of all positionable elements is equal to "block" for all positionable elements).
- All boxes have an absolute positioning scheme (i.e. the value of the "position" property is equal to "absolute" for all positionable elements).
- All boxes are contained within a single containing block defined by the root element. None of the boxes generated by the positionable elements contains another box.

The "layout" element determines how the elements in the document's body are positioned on

```
<smil>
  <head>
    <switch>
     <layout type="text/css">
        [region="r"] { top: 20px; left: 20px }
        #i2 { top: 30px; left: 30px }
     </layout>
     <layout>
       <region id="r" top="20" left="20" />
     </layout>
   </switch>
   </head>
   <body>
     <seq>
       <img region="r" src="http://www.w3.org/test" dur="10s" />
       <img id="i2" src="http://www.w3.org/test2" dur="5s" />
   </body>
<smil>
```

an abstract rendering surface (either visual or acoustic). If a document contains no layout element, the positioning of the body elements is implementation-dependent.

A SMIL document can contain alternative layout elements as content of a "switch" element (defined later). This can be used for example to describe the document's layout using different layout languages. The example in the figure shows how CSS2 can be used as alternative layout language to the SMIL basic layout language.

4.5 Alternate behavior Specification

Multimedia presentations are very demanding in terms of network bandwidth. Today, the network infrastructure offers no guarantees for the quality delivery of multimedia content. SMIL clients may access the documents from different locations with different network characteristics. Therefore multimedia documents must be adaptive.

```
<par>
  <text .../>
  <switch>
    <par system-screen-size="1280X1024" system-screen depth="16">
    ........
    </par>
    <par system-screen-size="640X480" system-screen-depth="32">
    ...
    </par>
    <par system-screen-size="640X480" system-screen-depth="16">
    ...
    </par>
  </switch>
</par>
...
```

The switch element allows an author to specify a set of alternative elements from which only one acceptable element should be selected. An element is acceptable if the element is a SMIL 1.0 element, the media-type can be decoded, and all of the test-attributes of the element evaluate to "true". In the previous example, the switch is based on the screen size and the screen depth or resolution. This allows the SMIL document to fit various environments while giving the authors the possibility to adapt the content accordingly.

In the second example, the switch is used to support multilingual documents. An audio object is available both in French and in English. Based on the user's preferred language, the player can choose one of these audio resources making the same document available for users understanding different languages.

```
...
<switch>
   <audio src="joe-audio-french" system-language="fr"/>
   <audio src="joe-audio-english" system-language="en"/>
</switch>
...
```

During the presentation, an element is selected as follows: the player evaluates the elements in the order in which they occur in the switch element. The first acceptable element is selected at the exclusion of all other elements within the switch.

Thus, authors should order the alternatives from the most desirable to the least desirable alternative. Furthermore, authors should place a relatively fail-safe alternative as the last item in the <switch> so that at least one item within the switch is chosen (unless this is explicitly not desired). Implementations should not arbitrarily pick an object within a <switch> when test-attributes for all fail. The reason behind this constraint is to keep different players behave in the same way in the same conditions.

4.6 Hypermedia linking in SMIL

Similarly to the Web infrastructure, one of the most important features of SMIL is its hyper-linking capability. This allows SMIL documents to contain references to other documents or parts of documents. The first important differences between links in HTML and those of SMIL are that the latter are time-dependent. For example, a piece of text can only be considered as actionable in the period of time where the anchored text is shown to the user. Furthermore, the destination or target of the link contains implicitly or explicitly a date when to start the target document. If not specified, the link traversal means the document should be started from its beginning instant. So both of the source and destination of links timing are to be taken into account.

When the user activates a given link, it is necessary to define its effect on the current presentation. In SMIL, three values of the *show* attribute define the way the link traversal must be handled by the client:

- **Replace:** the presentation of the SMIL document targeted by the destination link is started and replaces entirely the current presentation. This latter is ended and optionally added in the navigation stack of the SMIL Player.
- **New:** The presentation of the destination is started in a new window without affecting temporally or spatially the existing ones.
- **Pause:** The presentation of the destination is started and the presentation containing the source link is suspended. The presentation can be resumed either through control buttons or when the new presentation has ended.

The link activation within objects contained in SMIL documents, for example HTML documents, affects only those objects. So navigation within independent portions of SMIL document is preserved. For example, it is possible to include a VRML object and to make all the attached controls available to user without affecting the SMIL presentation. It is also possible to combine SMIL links with other object links by relating them spatially or temporally. In the following sections, we give some examples of such link uses.

4.6.1 Associating links with spatial subparts

In the following example, the screen space taken up by a video clip is split into two regions. A different link is associated with each of these regions. The activation time of the link

```
<video src="http://www.w3.org/CoolStuff">
  <anchor href="http://www.w3.org/AudioVideo" coords="0%,0%,50%,50%"/>
  <anchor href="http://www.w3.org/Style"      coords="50%,50%,100%,100%"/>
</video>
```

corresponds to the whole video duration. In this particular case, anchors inherit their duration from the enclosing element because only their spatial dimension is specified.

4.6.2 Associating links with temporal subparts

In the following example, the duration of a video clip is split into two subintervals. A different link is associated with each of these subintervals. Similarly to the previous case, the link is actionable on the whole surface of the video because the author has specified no particular

```
<video src="http://www.w3.org/CoolStuff">
  <anchor href="http://www.w3.org/AudioVideo" begin="0s" end="5s"/>
  <anchor href="http://www.w3.org/Style"      begin="5s" end="10s"/>
</video>
```

region. Anchors inherit in this case the bounding box of the video which may be shadowed by any possible stacking on top of the video.

4.6.3 Jumping to a subpart of an object

The following example contains a link from an element in one presentation A to the middle of a video object contained in another presentation B. This results in playing presentation B starting from second 5 in the video. The presentation would start as if the user had fast-forwarded the whole presentation to the point at which the designated fragment in the "CoolStuff" video begins.

```
Presentation A:
<a href="http://www.w3.org/mm/presentationB#tim">
   <video id="graph" src="rtsp://foo.com/graph.imf" region="l_window"/>
</a>
Presentation B:
<video src="http://www.w3.org/CoolStuff">
  <anchor id="joe" begin="0s" end="5s"/>
  <anchor id="tim" begin="5s" end="10s"/>
</video>
```

The semantic of the link that we have presented requires the player to restore the target document temporal structure and to jump directly to the point designated by the link. This requires also that some of the elements may have to be resumed from a mid-point in their presentation. This seems to be an easy task but a closer look at this problem shows that this may not be very obvious. In particular, when some of the objects have not a pre-defined duration, several cases can arise. A first solution to this problem is given in [Sab97].

4.6.4 Associating links with media-format internal identifiers

In the following example, two internal identifiers contained in the "CoolStuff" animation file are made accessible within SMIL. A link is associated with the objects designated by these identifiers. Identifies are markers within media objects which can be made available to the author

```
<ref src="http://www.w3.org/CoolStuff.anim">
  <anchor id="joe" fragment-id="AX789" href="http://www.cnn.com/joe"/>
  <anchor id="tim" fragment-id="AY994" href="http://www.w3.org/People/Berners-Lee"/>
</ref>
```

so that synchronization can be applied to a finer grain level. Examples can be found in scenarios where synchronization is applied between audio streams and text streams so that text elements appears only on the screen when they are laid out on the speakers. In future version, markers in media objects can be optionally contained in streams of anchors that are sent to the client in real time in parallel to the media objects.

4.6.5 Combining different uses of links

The following example shows how the different uses of associated links can be used in combination. In the presentation A, which is a simple video, the activation of the link leads to the second part of the video contained in the document B. The surface of the activation corresponds

```
Presentation A:
<a href="http://www.w3.org/mm/presentationB#tim">
  <video id="graph" src="rtsp://foo.com/graph.imf" region="l_window"/>
</a>
Presentation B:
<video src="http://www.w3.org/CoolStuff">
  <anchor id="joe" begin="0s" end="5s" coords="0%,0%,50%,50%"
          href="http://www.w3.org/"/>
  <anchor id="tim" begin="5s" end="10s" coords="0%,0%,50%,50%"
          href="http://www.w3.org/Tim"/>
</video>
```

to the region called *l_window* in the layout specification. This latter is referring in its top-left corner (given by the coordinates) to World Wide Web consortium Web site.

5 CONCLUSION

SMIL is a document format that eases both authoring and presentation of multimedia documents. It provides authors with an efficient and flexible way to specify a multimedia document content while retaining established declarative languages for temporal synchronization and spatial positioning. In this paper, we have presented the design principles of SMIL, which are based on a structured approach. We showed that temporal mark-up is an adequate concept for high-level document specification. It also provides the authors with the required transparency and abstractions. First, at the author's level, temporal relations allow the combination of intuitive graphic representations of scenarios like timelines which can be derived easily from the document structure. Second, SMIL supports a simple but powerful temporal model that makes document maintenance easier by allowing a nesting of sub-scenarios and preventing authors from introducing errors when writing complex presentations.

One of the main challenges of SMIL presentation engines is to dynamically adapt to the current presentation conditions. Despite the availability of a switch element which facilitates such adaptations, the presentation can be negatively affected by the bandwidth limitations, opaque objects (objects with no specified duration), unexpected delays and the system loads. Therefore, temporal constraints cannot be statically guaranteed. But we believe that quality of service through dynamic supervision can enhance the presentation through temporal prediction and constraints re-evaluation. Given the rich information the scenario contains, its predictive analysis can greatly facilitate the design of such an adaptive scheduling scheme. Information about the temporal ordering (sequential and parallel nodes) can be used in an observation and prediction loop: the information recorded at each observation instant can be used to predict what objects should be preloaded in a cache to minimize effects like jitter and skew.

Acknowledgments:

The author thanks all his colleagues of the Synchronized Multimedia Working Group at the World Wide Web Consortium who made this specification happen. This work was also supported by a grant from INRIA.

6 REFERENCES

[Kha95]

H. Khalfallah, A. Karmouch, ``An architecture and a data model for integrated multimedia documents and presentation applications", Multimedia Systems, vol. 1995, num. 3, pp. 238-250, 1995.

[Lay97]

N. Layaïda, "Madeus: système d'édition et de présentation de documents multimédia structurés". PhD Thesis, Universite Joseph Fourier, 12 June 1997.

[Lay96]

N. Layaïda et L. Sabry-Ismail, Maintaining Temporal Consistency of Multimedia Documents Using Constraint Networks, *International Conference on Multimedia Computing and Networking*, ed. M. Freeman, P. Jardetzky, H. M. Vin, Editors, San Jose, SPIE vol. 2667, pp. 124-135, USA, January 1996.

[SMI98]

SMIL : Synchronized Multimedia Integration Language. W3C Proposed Recommendation. World Wide Web Consortium. April 1998. See http://www.w3.org/AudioVideo/Group/WD-smil.

[Sab97]

L. Sabry-Ismail, C. Roisin, N. Layaïda, ``Navigation in Structured Multimedia Documents using Presentation Context", *Hypertextes et Hypermédias,* Hermès, ed. Paris, September 1997.

[CSS98]

Cascading Style Sheets, level 2", B. Bos, H. Lie, C. Lilley, I. Jacobs, 24 March 1998. Available at http://www.w3.org/TR/PR-CSS2/.

[XML98]

XML "Extensible Markup Language (XML) 1.0", T. Bray, J. Paoli, C.M. Sperberg-McQueen, editors, 10 February 1998. Available at http://www.w3.org/TR/REC-xml/

[RFC1808]

Relative Uniform Resource Locators, R. Fielding, June 1995. Available at ftp://ds.internic.net/rfc/rfc1808.txt.

[RFC2045]

Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies, N. Freed and N. Borenstein, November 1996. Available at ftp://ftp.internic.net/rfc/rfc2045.txt.

[SMPTE]

"Time and Control Codes for 24, 25 or 30 Frame-Per-Second Motion-Picture Systems - RP 136-1995". Society of Motion Picture & Television Engineers.

[HTM97]

"HTML 4.0 Specification", D. Raggett, A. Le Hors, I. Jacobs, 8 July 1997.
Available at http://www.w3.org/TR/REC-html40/.